Persistent Homology Tutorial

Xiaojin Zhu

Department of Computer Sciences
University of Wisconsin-Madison

jerryzhu@cs.wisc.edu

2013
Persistent homology

- A rapidly growing branch of topology
Persistent homology

- A rapidly growing branch of topology
- mathematically defines “holes” in data:
Persistent homology

- A rapidly growing branch of topology
- Mathematically defines “holes” in data:
  - $0^{th}$ order holes: clusters
Persistent homology

- A rapidly growing branch of topology
- mathematically defines “holes” in data:
  - $0^{th}$ order holes: clusters
  - $1^{st}$ order holes: holes

(Zhu, University of Wisconsin-Madison)
Persistent homology

- A rapidly growing branch of topology
- mathematically defines “holes” in data:
  - $0^{th}$ order holes: clusters
  - $1^{st}$ order holes: holes
  - $2^{nd}$ order holes: voids
Persistent homology

- A rapidly growing branch of topology
- Mathematically defines “holes” in data:
  - $0^{th}$ order holes: clusters
  - $1^{st}$ order holes: holes
  - $2^{nd}$ order holes: voids
  - Higher order holes, too

\[(\text{Zhu, University of Wisconsin-Madison})\]
Persistent homology

- A rapidly growing branch of topology
- mathematically defines “holes” in data:
  - $0^{th}$ order holes: clusters
  - $1^{st}$ order holes: holes
  - $2^{nd}$ order holes: voids
  - higher order holes, too
- Betti numbers: the number of $k^{th}$ order holes
Betti number examples

(1,0,0,0,...)  (1,1,0,0,...)  (1,2,1,0,...)  (1,2,1,0,...)  (1,0,1,0,...)

[Reproduced from Singh et al. J. Vision 2008]
Plan of this talk

- Persistent homology tutorial
Plan of this talk

- Persistent homology tutorial
- An application in natural language processing
Holes and equivalent rubber bands

- blue $\sim$ green, not red
Holes and equivalent rubber bands

- blue $\sim$ green, not red
- two equivalent classes $\Leftrightarrow$ one hole.
Group Theory

Definition

A group \( \langle G, * \rangle \) is a set \( G \) with a binary operation \( * \) such that

1. (associative) \( a * (b * c) = (a * b) * c \) for all \( a, b, c \in G \).
2. (identity) \( \exists e \in G \) so that \( e * a = a * e = a \) for all \( a \in G \).
3. (inverse) \( \forall a \in G, \exists a' \in G \) where \( a * a' = a' * a = e \).

Examples: \( \langle \mathbb{Z}, + \rangle \), \( \langle \mathbb{R}, + \rangle \), \( \langle \mathbb{R} \{0\}, \times \rangle \), \( \langle \mathbb{Z}_2, + \rangle \).

All our groups \( G \) are abelian: \( \forall a, b \in G, a * b = b * a \).

(Zhu, University of Wisconsin-Madison)
Group Theory

Definition

A group \( \langle G, * \rangle \) is a set \( G \) with a binary operation \( * \) such that

1. **(associative)** \( a * (b * c) = (a * b) * c \) for all \( a, b, c \in G \).
Group Theory

Definition

A group $\langle G, \ast \rangle$ is a set $G$ with a binary operation $\ast$ such that

1. **(associative)** $a \ast (b \ast c) = (a \ast b) \ast c$ for all $a, b, c \in G$.

2. **(identity)** $\exists e \in G$ so that $e \ast a = a \ast e = a$ for all $a \in G$. 

Examples: $\langle \mathbb{Z}, + \rangle$, $\langle \mathbb{R}, + \rangle$, $\langle \mathbb{R}^+ \setminus \{0\}, \times \rangle$, $\langle \mathbb{Z}_2, + \rangle$. 

All our groups $G$ are abelian: $\forall a, b \in G, a \ast b = b \ast a$.
Definition

A group \( \langle G, \ast \rangle \) is a set \( G \) with a binary operation \( \ast \) such that

1. (associative) \( a \ast (b \ast c) = (a \ast b) \ast c \) for all \( a, b, c \in G \).
2. (identity) \( \exists e \in G \) so that \( e \ast a = a \ast e = a \) for all \( a \in G \).
3. (inverse) \( \forall a \in G, \exists a' \in G \) where \( a \ast a' = a' \ast a = e \).
A group \( \langle G, \ast \rangle \) is a set \( G \) with a binary operation \( \ast \) such that

1. **(associative)** \( a \ast (b \ast c) = (a \ast b) \ast c \) for all \( a, b, c \in G \).
2. **(identity)** \( \exists e \in G \) so that \( e \ast a = a \ast e = a \) for all \( a \in G \).
3. **(inverse)** \( \forall a \in G, \exists a' \in G \) where \( a \ast a' = a' \ast a = e \).

Examples: \( \langle \mathbb{Z}, + \rangle, \langle \mathbb{R}, + \rangle, \langle \mathbb{R}_+, \times \rangle, \langle \mathbb{R}\backslash\{0\}, \times \rangle \).
Definition

A group $\langle G, \ast \rangle$ is a set $G$ with a binary operation $\ast$ such that

1. (associative) $a \ast (b \ast c) = (a \ast b) \ast c$ for all $a, b, c \in G$.
2. (identity) $\exists e \in G$ so that $e \ast a = a \ast e = a$ for all $a \in G$.
3. (inverse) $\forall a \in G, \exists a' \in G$ where $a \ast a' = a' \ast a = e$.

Examples: $\langle \mathbb{Z}, + \rangle$, $\langle \mathbb{R}, + \rangle$, $\langle \mathbb{R}_+, \times \rangle$, $\langle \mathbb{R} \setminus \{0\}, \times \rangle$.

$\mathbb{Z}_2$

\[
\begin{array}{c|cc}
+ & 0 & 1 \\
\hline
0 & 0 & 1 \\
1 & 1 & 0 \\
\end{array}
\]
**Group Theory**

**Definition**

A **group** \( \langle G, * \rangle \) is a set \( G \) with a **binary operation** \( * \) such that

1. **(associative)** \( a * (b * c) = (a * b) * c \) for all \( a, b, c \in G \).
2. **(identity)** \( \exists e \in G \) so that \( e * a = a * e = a \) for all \( a \in G \).
3. **(inverse)** \( \forall a \in G, \exists a' \in G \) where \( a * a' = a' * a = e \).

- Examples: \( \langle \mathbb{Z}, + \rangle \), \( \langle \mathbb{R}, + \rangle \), \( \langle \mathbb{R}_+, \times \rangle \), \( \langle \mathbb{R} \setminus \{0\}, \times \rangle \).
- \( \mathbb{Z}_2 \)

\[
\begin{array}{c|cc}
+2 & 0 & 1 \\
\hline
0 & 0 & 1 \\
1 & 1 & 0 \\
\end{array}
\]

- All our groups \( G \) are **abelian**: \( \forall a, b \in G, a * b = b * a \).
A subset $H \subseteq G$ of a group $\langle G, \ast \rangle$ is a **subgroup** of $G$ if $\langle H, \ast \rangle$ is itself a group.
Subgroup

Definition

A subset $H \subseteq G$ of a group $\langle G, \ast \rangle$ is a subgroup of $G$ if $\langle H, \ast \rangle$ is itself a group.

- $\{e\}$ is the trivial subgroup of any group $G$
Subgroup

**Definition**

A subset \( H \subseteq G \) of a group \( \langle G, * \rangle \) is a **subgroup** of \( G \) if \( \langle H, * \rangle \) is itself a group.

- \( \{e\} \) is the trivial subgroup of any group \( G \)
- \( \langle \mathbb{R}_+, \times \rangle \) is a subgroup of \( \langle \mathbb{R} \setminus \{0\}, \times \rangle \)
Subgroup

Definition

A subset $H \subseteq G$ of a group $\langle G, * \rangle$ is a subgroup of $G$ if $\langle H, * \rangle$ is itself a group.

- $\{e\}$ is the trivial subgroup of any group $G$
- $\langle \mathbb{R}_+, \times \rangle$ is a subgroup of $\langle \mathbb{R}\setminus\{0\}, \times \rangle$
- not $\langle \mathbb{R}_-, \times \rangle$
Given a subgroup $H$ of an abelian group $G$, for any $a \in G$, the set $a \ast H = \{a \ast h \mid h \in H\}$ is the coset of $H$ represented by $a$. 

For example, $R = \mathbb{R}$ and $G = \mathbb{Z}$, $3 \times R +$ is a coset which is the same as $R + -1 \times R = R -$ is another coset (not a subgroup). Cosets have equal sizes and partition $G$. 

(Zhu, University of Wisconsin-Madison)
Coset

**Definition**

Given a subgroup $H$ of an abelian group $G$, for any $a \in G$, the set $a \ast H = \{a \ast h \mid h \in H\}$ is the coset of $H$ represented by $a$.

- $H = \mathbb{R}_+, \ G = \mathbb{R}\{0\}$
Coset

Definition

Given a subgroup $H$ of an abelian group $G$, for any $a \in G$, the set $a \cdot H = \{ a \cdot h \mid h \in H \}$ is the coset of $H$ represented by $a$.

- $H = \mathbb{R}_+$, $G = \mathbb{R} \setminus \{0\}$
- $3.14 \times \mathbb{R}_+$ is a coset which is the same as $\mathbb{R}_+$
Coset

Definition

Given a subgroup $H$ of an abelian group $G$, for any $a \in G$, the set $a \cdot H = \{ a \cdot h \mid h \in H \}$ is the coset of $H$ represented by $a$.

- $H = \mathbb{R}_+$, $G = \mathbb{R}\setminus\{0\}$
- $3.14 \times \mathbb{R}_+$ is a coset which is the same as $\mathbb{R}_+$
- $-1 \times \mathbb{R}_+ = \mathbb{R}_-$ is another coset (not a subgroup)
Coset

Definition

Given a subgroup $H$ of an abelian group $G$, for any $a \in G$, the set $a \ast H = \{a \ast h \mid h \in H\}$ is the coset of $H$ represented by $a$.

- $H = \mathbb{R}_+, \ G = \mathbb{R}\{0\}$
- $3.14 \times \mathbb{R}_+$ is a coset which is the same as $\mathbb{R}_+$
- $-1 \times \mathbb{R}_+ = \mathbb{R}_-$ is another coset (not a subgroup)
- cosets have equal sizes and partition $G$. 

(Zhu, University of Wisconsin-Madison)
Homomorphism

Definition

A map $\phi : G \mapsto G'$ is a **homomorphism** if $\phi(a \ast b) = \phi(a) \ast \phi(b)$ for $\forall a, b \in G$. 

(Definition excerpted from Zhu, University of Wisconsin-Madison)
Definition

A map $\phi : G \mapsto G'$ is a **homomorphism** if $\phi(a \ast b) = \phi(a) \ast \phi(b)$ for $\forall a, b \in G$.

- $\langle \mathbb{R}_+, \times \rangle$ to $\langle \mathbb{Z}_2, +_2 \rangle$: trivial homomorphism $\phi(a) = 0$, $\forall a \in \mathbb{R}_+$
A map $\phi : G \mapsto G'$ is a homomorphism if $\phi(a \ast b) = \phi(a) \ast \phi(b)$ for all $a, b \in G$.

- $\langle \mathbb{R}_+, \times \rangle$ to $\langle \mathbb{Z}_2, +_2 \rangle$: trivial homomorphism $\phi(a) = 0$, $\forall a \in \mathbb{R}_+$
- Negation in natural language: $G_N$

<table>
<thead>
<tr>
<th></th>
<th>$\square$</th>
<th>not</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\times$</td>
<td>$\square$</td>
<td>not</td>
</tr>
<tr>
<td>not</td>
<td>not</td>
<td>$\square$</td>
</tr>
</tbody>
</table>

Homomorphism (isomorphism) from $G_N$ to $\mathbb{Z}_2$: $\phi(\square) = 0$, $\phi($not$) = 1$. 

(Zhu, University of Wisconsin-Madison)
The kernel of a homomorphism $\phi : G \rightarrow G'$ is
\[ \ker \phi = \{ a \in G \mid \phi(a) = e' \} . \]
Kernel

Definition

The kernel of a homomorphism \( \phi : G \mapsto G' \) is
\[
\ker \phi = \{ a \in G \mid \phi(a) = e' \}.
\]

- In the \( \phi : G_N \mapsto \mathbb{Z}_2 \) example, \( \ker \phi = \{ \square \} \).
Kernel

Definition

The kernel of a homomorphism $\phi : G \mapsto G'$ is 

$\ker \phi = \{ a \in G \mid \phi(a) = e' \}$.

- In the $\phi : G_N \mapsto \mathbb{Z}_2$ example, $\ker \phi = \{ \square \}$.
- Another example: $\phi : \langle \mathbb{R} \setminus \{0\}, \times \rangle \mapsto G_N$ by $\phi(a) = \square$ if $a > 0$ and “not” if $a < 0$. $\ker \phi = \mathbb{R}_+$
Kernel

**Definition**

*The kernel of a homomorphism* $\phi : G \mapsto G'$ *is*

$$\ker \phi = \{ a \in G \mid \phi(a) = e' \}.$$  

- In the $\phi : G_N \mapsto \mathbb{Z}_2$ example, $\ker \phi = \{ \square \}$.
- Another example: $\phi : \langle \mathbb{R} \setminus \{0\}, \times \rangle \mapsto G_N$ by $\phi(a) = \square$ if $a > 0$ and “not” if $a < 0$. $\ker \phi = \mathbb{R}_+$
- For any homomorphism $\phi : G \mapsto G'$, $\ker \phi$ is a subgroup of $G$.  

(Zhu, University of Wisconsin-Madison)

Persistent homology
**Definition**

The kernel of a homomorphism $\phi : G \to G'$ is 
\[\ker \phi = \{a \in G \mid \phi(a) = e'\} .\]

- In the $\phi : G_N \to \mathbb{Z}_2$ example, $\ker \phi = \{\square\}$.
- Another example: $\phi : \langle \mathbb{R}\setminus\{0\}, \times \rangle \to G_N$ by $\phi(a) = \square$ if $a > 0$ and “not” if $a < 0$. $\ker \phi = \mathbb{R}_+$.
- For any homomorphism $\phi : G \to G'$, $\ker \phi$ is a subgroup of $G$.
- Cosets $a \ast \ker \phi$ partition $G$. 

(Zhu, University of Wisconsin-Madison)
Quotient group

Let \( \langle H, * \rangle \) be a subgroup of an abelian group \( \langle G, * \rangle \).
Quotient group

- Let \( \langle H, \ast \rangle \) be a subgroup of an abelian group \( \langle G, \ast \rangle \).
- A new operation on the cosets of \( H \): 
  \[ (a \ast H) \ast (b \ast H) = (a \ast b) \ast H, \forall a, b \in G. \]
Quotient group

- Let $\langle H, * \rangle$ be a subgroup of an abelian group $\langle G, * \rangle$.
- A new operation on the cosets of $H$:
  \[(a * H) \star (b * H) = (a * b) * H, \forall a, b \in G.\]

**Definition**

*The cosets $\{a * H \mid a \in G\}$ under the operation $\star$ form a group, called the quotient group $G/H$.***
Quotient group

- Let $\langle H, * \rangle$ be a subgroup of an abelian group $\langle G, * \rangle$.
- A new operation on the cosets of $H$:
  \[(a * H) * (b * H) = (a * b) * H, \forall a, b \in G.\]

**Definition**

The cosets $\{a * H \mid a \in G\}$ under the operation $*$ form a group, called the quotient group $G/H$.

- Example: $G = \mathbb{R}\{0\}$ and $\ker \phi = \mathbb{R}_+$, two cosets: $\mathbb{R}_+$ and $\mathbb{R}_-$. 
Quotient group

- Let $\langle H, \star \rangle$ be a subgroup of an abelian group $\langle G, \star \rangle$.
- A new operation on the cosets of $H$:
  $$(a \star H) \star (b \star H) = (a \star b) \star H, \forall a, b \in G.$$ 

**Definition**

The cosets $\{a \star H \mid a \in G\}$ under the operation $\star$ form a group, called the quotient group $G/H$.

- Example: $G = \mathbb{R} \setminus \{0\}$ and $\ker \phi = \mathbb{R}_+$, two cosets: $\mathbb{R}_+$ and $\mathbb{R}_-$. 
- The quotient group $(\mathbb{R} \setminus \{0\})/\mathbb{R}_+$ has the two coset elements.
Let \( \langle H, \ast \rangle \) be a subgroup of an abelian group \( \langle G, \ast \rangle \).

A new operation on the cosets of \( H \):
\[
(a \ast H) \ast (b \ast H) = (a \ast b) \ast H, \forall a, b \in G.
\]

**Definition**

The cosets \( \{a \ast H \mid a \in G\} \) under the operation \( \ast \) form a group, called the quotient group \( G/H \).

**Example:** \( G = \mathbb{R}\setminus\{0\} \) and \( \ker \phi = \mathbb{R}_+ \), two cosets: \( \mathbb{R}_+ \) and \( \mathbb{R}_- \).

The quotient group \( (\mathbb{R}\setminus\{0\})/\mathbb{R}_+ \) has the two coset elements.

\[
\mathbb{R}_- \ast \mathbb{R}_- = (-1 \times \mathbb{R}_+) \ast (-1 \times \mathbb{R}_+) = (-1 \times -1) \times \mathbb{R}_+ = 1 \times \mathbb{R}_+ = \mathbb{R}_+.
\]
Quotient group

- Let \( \langle H, \ast \rangle \) be a subgroup of an abelian group \( \langle G, \ast \rangle \).
- A new operation on the cosets of \( H \):
  \[
  (a \ast H) \ast (b \ast H) = (a \ast b) \ast H, \forall a, b \in G.
  \]

Definition

The cosets \( \{a \ast H \mid a \in G\} \) under the operation \( \ast \) form a group, called the quotient group \( G/H \).

Example: \( G = \mathbb{R}\setminus\{0\} \) and \( \ker\phi = \mathbb{R}_+ \), two cosets: \( \mathbb{R}_+ \) and \( \mathbb{R}_- \).
- The quotient group \( (\mathbb{R}\setminus\{0\})/\mathbb{R}_+ \) has the two coset elements.
- \( \mathbb{R}_- \ast \mathbb{R}_- = (-1 \times \mathbb{R}_+) \ast (-1 \times \mathbb{R}_+) = (-1 \times -1) \times \mathbb{R}_+ = 1 \times \mathbb{R}_+ = \mathbb{R}_+ \).
- This quotient group \( (\mathbb{R}\setminus\{0\})/\mathbb{R}_+ \) is isomorphic to \( \mathbb{Z}_2 \).
Rank

**Definition**

Let $S$ be a subset of a group $G$. The *subgroup generated by* $S$, $\langle S \rangle$, is the subgroup of all elements of $G$ that can expressed as the finite operation of elements in $S$ and their inverses.

$\mathbb{Z}$ is itself the subgroup generated by $\{1\}$.

$\mathbb{Z} \times \mathbb{Z}$ is the subgroup generated by $\{(0, 1), (1, 0)\}$.

$\text{rank}(G)$ is the size of the smallest subset that generates $G$.

$\text{rank}(\mathbb{Z}) = 1$ since $\mathbb{Z} = \langle \{1\} \rangle$.

$\text{rank}(\mathbb{Z} \times \mathbb{Z}) = 2$ since $\mathbb{Z} \times \mathbb{Z} = \langle \{(0, 1), (1, 0)\} \rangle$. 

(Zhu, University of Wisconsin-Madison)
**Rank**

**Definition**

Let $S$ be a subset of a group $G$. The **subgroup generated by $S$, $\langle S \rangle$, is the subgroup of all elements of $G$ that can expressed as the finite operation of elements in $S$ and their inverses.**

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
Rank

Definition

Let $S$ be a subset of a group $G$. The subgroup generated by $S$, $\langle S \rangle$, is the subgroup of all elements of $G$ that can expressed as the finite operation of elements in $S$ and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$. 

(Zhu, University of Wisconsin-Madison)
Rank

Definition
Let $S$ be a subset of a group $G$. The subgroup generated by $S$, $\langle S \rangle$, is the subgroup of all elements of $G$ that can expressed as the finite operation of elements in $S$ and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$.

Definition
The rank of a group $G$ is $\text{rank}(G) = \min\{|S| \mid S \subseteq G, \langle S \rangle = G\}$. 
**Rank**

**Definition**

Let $S$ be a subset of a group $G$. The subgroup generated by $S$, $\langle S \rangle$, is the subgroup of all elements of $G$ that can expressed as the finite operation of elements in $S$ and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$.

**Definition**

The rank of a group $G$ is $\text{rank}(G) = \min\{|S| \mid S \subseteq G, \langle S \rangle = G \}$.

- $\text{rank}(G)$ is the size of the smallest subset that generates $G$. 

(Zhu, University of Wisconsin-Madison)
Rank

Definition

Let $S$ be a subset of a group $G$. The subgroup generated by $S$, $\langle S \rangle$, is the subgroup of all elements of $G$ that can expressed as the finite operation of elements in $S$ and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$.

Definition

The rank of a group $G$ is $\text{rank}(G) = \min\{|S| \mid S \subseteq G, \langle S \rangle = G\}$.

- $\text{rank}(G)$ is the size of the smallest subset that generates $G$.
- $\text{rank}(\mathbb{Z}) = 1$ since $\mathbb{Z} = \langle \{1\} \rangle$. 
Rank

**Definition**

Let $S$ be a subset of a group $G$. The **subgroup generated by** $S$, $\langle S \rangle$, is the subgroup of all elements of $G$ that can expressed as the finite operation of elements in $S$ and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$.

**Definition**

The **rank of a group** $G$ is $\text{rank}(G) = \min\{|S| \mid S \subseteq G, \langle S \rangle = G\}$.

- $\text{rank}(G)$ is the size of the smallest subset that generates $G$.
- $\text{rank}(\mathbb{Z}) = 1$ since $\mathbb{Z} = \langle \{1\} \rangle$.
- $\text{rank}(\mathbb{Z} \times \mathbb{Z}) = 2$ since $\mathbb{Z} \times \mathbb{Z} = \langle \{(0, 1), (1, 0)\} \rangle$. 
The group of rubber bands

- To count “holes” in homology, consider the group of cycles (the rubber bands)

Computation: need discrete rubber bands $\Rightarrow$ simplicial complex
The group of rubber bands

- To count “holes” in homology, consider the group of cycles (the rubber bands)
- The kernel: “uninteresting rubber bands” that do not surround holes
The group of rubber bands

- To count “holes” in homology, consider the group of cycles (the rubber bands)
- The kernel: “uninteresting rubber bands” that do not surround holes
- The quotient group “all rubber bands” / “uninteresting rubber bands” will identify holes.
The group of rubber bands

- To count “holes” in homology, consider the group of cycles (the rubber bands)
- The kernel: “uninteresting rubber bands” that do not surround holes
- The quotient group “all rubber bands” / “uninteresting rubber bands” will identify holes.
- Computation: need discrete rubber bands $\Rightarrow$ simplicial complex
Simplex

Definition

A \textit{p-simplex} \( \sigma \) is the convex hull of \( p + 1 \) affinely independent points \( x_0, x_1, \ldots, x_p \in \mathbb{R}^d \). We denote \( \sigma = \text{conv}\{x_0, \ldots, x_p\} \). \textbf{The dimension of} \( \sigma \) \textbf{is} \( p \).
Simplex

Definition

A $p$-simplex $\sigma$ is the convex hull of $p + 1$ affinely independent points $x_0, x_1, \ldots, x_p \in \mathbb{R}^d$. We denote $\sigma = \text{conv}\{x_0, \ldots, x_p\}$. The dimension of $\sigma$ is $p$.

- $p = 0, 1, 2, 3$
Simplicial complex

Definition

A simplicial complex $K$ is a finite collection of simplices such that $\sigma \in K$ and $\tau$ being a face of $\sigma$ implies $\tau \in K$, and $\sigma, \sigma' \in K$ implies $\sigma \cap \sigma'$ is either empty or a face of both $\sigma$ and $\sigma'$. 
Simplicial complex

**Definition**

A *simplicial complex* $K$ is a finite collection of simplices such that $\sigma \in K$ and $\tau$ being a face of $\sigma$ implies $\tau \in K$, and $\sigma, \sigma' \in K$ implies $\sigma \cap \sigma'$ is either empty or a face of both $\sigma$ and $\sigma'$. 

- Properly aligned
A simplicial complex $K$ is a finite collection of simplices such that $\sigma \in K$ and $\tau$ being a face of $\sigma$ implies $\tau \in K$, and $\sigma, \sigma' \in K$ implies $\sigma \cap \sigma'$ is either empty or a face of both $\sigma$ and $\sigma'$. 

Properly aligned

Simplicial complex $=$ the yellow space in the rubber band picture
Chain

**Definition**

A $p$-chain is a subset of $p$-simplices in a simplicial complex $K$.

Example: $K = \text{tetrahedron}$. A 2-chain is a subset of the four triangles. 2 4 2 distinct 2-chains. 2 6 2 distinct 1-chains (subsets of edges).

A $p$-chain does not have to be connected.
Definition

A $p$-chain is a subset of $p$-simplices in a simplicial complex $K$.

- Example: $K=$tetrahedron.
Chain

Definition

A $p$-chain is a subset of $p$-simplices in a simplicial complex $K$.

- Example: $K=$tetrahedron.
- A 2-chain is a subset of the four triangles.
**Chain**

**Definition**

A $p$-chain is a subset of $p$-simplices in a simplicial complex $K$.

- Example: $K$ = tetrahedron.
- A 2-chain is a subset of the four triangles.
- $2^4$ distinct 2-chains.
Chain

Definition

A $p$-chain is a subset of $p$-simplices in a simplicial complex $K$.

- Example: $K$=tetrahedron.
- A 2-chain is a subset of the four triangles.
- $2^4$ distinct 2-chains.
- $2^6$ distinct 1-chains (subsets of edges).
A \( p \)-chain is a subset of \( p \)-simplices in a simplicial complex \( K \).

- Example: \( K \) = tetrahedron.
- A 2-chain is a subset of the four triangles.
- \( 2^4 \) distinct 2-chains.
- \( 2^6 \) distinct 1-chains (subsets of edges).
- Left: a 2-chain, right: a 1-chain
Definition

A $p$-chain is a subset of $p$-simplices in a simplicial complex $K$.

- Example: $K=$tetrahedron.
- A 2-chain is a subset of the four triangles.
- $2^4$ distinct 2-chains.
- $2^6$ distinct 1-chains (subsets of edges).
- Left: a 2-chain, right: a 1-chain

- A $p$-chain does not have to be connected.
Chain group

**Definition**

*The set of $p$-chains of a simplicial complex $K$ form a $p$-chain group $C_p$.***
Chain group

Definition

The set of $p$-chains of a simplicial complex $K$ form a $p$-chain group $C_p$.

- Mod-2 addition
Boundary

Definition

The boundary of a $p$-simplex is the set of $(p - 1)$-simplices faces.
The boundary of a $p$-simplex is the set of $(p - 1)$-simplices faces.

- boundary of a tetrahedron = the four triangles faces
Boundary

Definition

The **boundary** of a $p$-simplex is the set of $(p - 1)$-simplices faces.

- boundary of a tetrahedron = the four triangles faces
- boundary of a triangle = the three edges
Boundary

Definition

The **boundary** of a $p$-simplex is the set of $(p - 1)$-simplices faces.

- boundary of a tetrahedron = the four triangles faces
- boundary of a triangle = the three edges
- boundary of an edge = its two vertices
Definition

The boundary of a \( p \)-chain is the Mod-2 sum of the boundaries of its simplices. Taking the boundary is a group homomorphism \( \partial_p \) from \( C_p \) to \( C_{p-1} \).
**Boundary of a $p$-chain**

**Definition**

*The boundary of a $p$-chain is the Mod-2 sum of the boundaries of its simplices. Taking the boundary is a group homomorphism $\partial_p$ from $C_p$ to $C_{p-1}$.***

- Faces shared by an even number of $p$-simplices in the chain will cancel out:

![Diagram of boundary operation]
Definition

A \( p \text{-cycle} \) \( c \) is a \( p \text{-chain with empty boundary: } \partial_p c = 0 \) (the identity in \( C_{p-1} \)).
Cycles

Definition

A \( p \)-cycle \( c \) is a \( p \)-chain with empty boundary: \( \partial_p c = 0 \) (the identity in \( C_{p-1} \)).

- Discrete \( p \)-dimensional “rubber bands”
Cycles

**Definition**

A \( p \)-cycle \( c \) is a \( p \)-chain with empty boundary: \( \partial_p c = 0 \) (the identity in \( C_{p-1} \)).

- Discrete \( p \)-dimensional “rubber bands”
- Left: a 1-cycle; Right: not a cycle
A \textit{p-cycle} $c$ is a \textit{p-chain} with empty boundary: $\partial_p c = 0$ (the identity in $C_{p-1}$).

- Discrete $p$-dimensional “rubber bands”
- Left: a 1-cycle; Right: not a cycle
- $Z_p = \text{all } p\text{-cycles} (\text{all rubber bands})$
Cycles

**Definition**

A $p$-cycle $c$ is a $p$-chain with empty boundary: $\partial_pc = 0$ (the identity in $C_{p-1}$).

- Discrete $p$-dimensional “rubber bands”
- Left: a 1-cycle; Right: not a cycle

$Z_p = $ all $p$-cycles (all rubber bands)

$\partial_pZ_p = 0$: $Z_p$ is the kernel $\ker\partial_p$ and a subgroup of $C_p$. 
The boundary of any \((p + 1)\)-chain is always a \(p\)-cycle.

\begin{itemize}
  \item \(C_1\)
  \item \(C_2\)
  \item \(C_3\)
\end{itemize}
Boundary-Cycle

- The boundary of any \((p + 1)\)-chain is always a \(p\)-cycles

\[
\begin{array}{ccc}
\text{c}_1 & \text{c}_2 & \text{c}_3 \\
\end{array}
\]

Definition

A \(p\)-boundary-cycle is a \(p\)-cycle that is also the boundary of some \((p + 1)\)-chain.
The boundary of any \((p + 1)\)-chain is always a \(p\)-cycles.

\[ B_p = \partial_{p+1} C_{p+1}, \text{ the } p\text{-boundary-cycles.} \]
Boundary-Cycle

- The boundary of any \((p + 1)\)-chain is always a \(p\)-cycles

\[ C_1 \quad C_2 \quad C_3 \]

Definition

A \(p\)-boundary-cycle is a \(p\)-cycle that is also the boundary of some \((p + 1)\)-chain.

- Let \(B_p = \partial_{p+1} C_{p+1}\), the \(p\)-boundary-cycles.
- \(B_p\) are the uninteresting rubber bands (e.g., \(B_1 = \{0, c_1\}\))
The boundary of any $(p + 1)$-chain is always a $p$-cycles.

**Definition**

A $p$-boundary-cycle is a $p$-cycle that is also the boundary of some $(p + 1)$-chain.

- Let $B_p = \partial_{p+1} C_{p+1}$, the $p$-boundary-cycles.
- $B_p$ are the uninteresting rubber bands (e.g., $B_1 = \{0, c_1\}$)
- $B_p$ is a subgroup of $Z_p$ (all rubber bands).
Interesting rubber bands

- $c_2$ and $c_3$ in $Z_1$ but not in $B_1$
Interesting rubber bands

- $c_2$ and $c_3$ in $Z_1$ but not in $B_1$
- We can drag rubber band $c_2$ over the yellow triangle to make $c_3$
Interesting rubber bands

- $c_2$ and $c_3$ in $Z_1$ but not in $B_1$
- We can drag rubber band $c_2$ over the yellow triangle to make $c_3$
- Formally, $c_3 = c_2 + c_1$. 

$C_1$ $C_2$ $C_3$
Interesting rubber bands

- $c_2$ and $c_3$ in $Z_1$ but not in $B_1$
- We can drag rubber band $c_2$ over the yellow triangle to make $c_3$
- Formally, $c_3 = c_2 + c_1$.
- $c_2$ and $c_3$ are equivalent in the hole they surround.
Interesting rubber bands

- \( c_2 \) and \( c_3 \) in \( Z_1 \) but not in \( B_1 \)
- We can drag rubber band \( c_2 \) over the yellow triangle to make \( c_3 \)
- Formally, \( c_3 = c_2 + c_1 \).
- \( c_2 \) and \( c_3 \) are equivalent in the hole they surround.
- The equivalence class: \( c + B_p \)
Homology group

Definition

The $p$-th homology group is the quotient group $H_p = \mathbb{Z}_p / B_p$. 

Example:

All the 1-cycles: $\mathbb{Z}_1 = \{0, c_1, c_2, c_3\}$.

The uninteresting 1-cycles: $B_1 = \{0, c_1\}$, a subgroup of $\mathbb{Z}_1$.

The interesting 1-cycles: $c_2 + B_1 = c_3 + B_1 = \{c_2, c_3\}$.

The homology group $H_1 = \mathbb{Z}_1 / B_1$ isomorphic to $\mathbb{Z}_2$. (Zhu, University of Wisconsin-Madison)
Homology group

**Definition**

The \( p \)-th homology group is the quotient group \( H_p = \mathbb{Z}_p / B_p \).

- **Example:**

![Diagram of cycles](image)
Homology group

Definition

The \( p \)-th homology group is the quotient group \( H_p = \mathbb{Z}_p / B_p \).

- Example:

- All the 1-cycles: \( Z_1 = \{0, c_1, c_2, c_3\} \).
Homology group

Definition

The $p$-th homology group is the quotient group $H_p = Z_p / B_p$.

- Example:

- All the 1-cycles: $Z_1 = \{0, c_1, c_2, c_3\}$.
- The uninteresting 1-cycles: $B_1 = \{0, c_1\}$, a subgroup of $Z_1$. 

(Zhu, University of Wisconsin-Madison)
Homology group

**Definition**

The $p$-th homology group is the quotient group $H_p = Z_p / B_p$.

- **Example:**

  - All the 1-cycles: $Z_1 = \{0, c_1, c_2, c_3\}$.
  - The uninteresting 1-cycles: $B_1 = \{0, c_1\}$, a subgroup of $Z_1$.
  - The interesting 1-cycles: $c_2 + B_1 = c_3 + B_1 = \{c_2, c_3\}$
Homology group

**Definition**

The \( p \)-th homology group is the quotient group \( H_p = Z_p/B_p \).

- **Example:**
  
  \[
  \text{All the 1-cycles : } Z_1 = \{0, c_1, c_2, c_3\}.
  \]
  
  \[
  \text{The uninteresting 1-cycles: } B_1 = \{0, c_1\}, \text{ a subgroup of } Z_1.
  \]
  
  \[
  \text{The interesting 1-cycles: } c_2 + B_1 = c_3 + B_1 = \{c_2, c_3\}
  \]
  
  \[
  \text{The homology group } H_1 = Z_1/B_1 \text{ isomorphic to } \mathbb{Z}_2
  \]
Betti number

Definition

The $p$-th Betti number is the rank of the homology group: $\beta_p = \text{rank}(H_p)$. 

In our example, $\beta_1 = \text{rank}(\mathbb{Z}^2) = 1$ (one 1st-order hole). $\beta_p$ is the number of independent $p$-th holes. A tetrahedron has $\beta_0 = 1$ (connected), $\beta_1 = \beta_2 = 0$ (no holes or voids). A hollow tetrahedron has $\beta_0 = 1$, $\beta_1 = 0$, $\beta_2 = 1$. Removing the four triangle faces, the edge skeleton has $\beta_0 = 1$, $\beta_1 = 3$ (one is the sum of the other three), $\beta_2 = 0$ (no more void). Removing the edges, $\beta_0 = 4$ (4 vertices) and $\beta_1 = \beta_2 = 0$. 

(Zhu, University of Wisconsin-Madison)
Betti number

Definition

The \( p \)-th Betti number is the rank of the homology group: \( \beta_p = \text{rank}(H_p) \).

- In our example, \( \beta_1 = \text{rank}(\mathbb{Z}_2) = 1 \) (one 1st-order hole)
Betti number

Definition

The $p$-th Betti number is the rank of the homology group: $\beta_p = \text{rank}(H_p)$.

- In our example, $\beta_1 = \text{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- $\beta_p$ is the number of independent $p$-th holes.
Betti number

Definition

The $p$-th Betti number is the rank of the homology group: $\beta_p = \text{rank}(H_p)$.

- In our example, $\beta_1 = \text{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- $\beta_p$ is the number of independent $p$-th holes.
- A tetrahedron has $\beta_0 = 1$ (connected), $\beta_1 = \beta_2 = 0$ (no holes or voids)
Betti number

**Definition**

The $p$-th **Betti number** is the rank of the homology group: $\beta_p = \text{rank}(H_p)$.

- In our example, $\beta_1 = \text{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- $\beta_p$ is the number of independent $p$-th holes.
- A tetrahedron has $\beta_0 = 1$ (connected), $\beta_1 = \beta_2 = 0$ (no holes or voids)
- A **hollow** tetrahedron has $\beta_0 = 1, \beta_1 = 0, \beta_2 = 1
Betti number

Definition

The $p$-th Betti number is the rank of the homology group: $\beta_p = \text{rank}(H_p)$.

- In our example, $\beta_1 = \text{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- $\beta_p$ is the number of independent $p$-th holes.
- A tetrahedron has $\beta_0 = 1$ (connected), $\beta_1 = \beta_2 = 0$ (no holes or voids)
- A hollow tetrahedron has $\beta_0 = 1$, $\beta_1 = 0$, $\beta_2 = 1$
- Removing the four triangle faces, the edge skeleton has $\beta_0 = 1$, $\beta_1 = 3$ (one is the sum of the other three), $\beta_2 = 0$ (no more void).
Betti number

**Definition**

The *p*-th **Betti number** is the rank of the homology group: $\beta_p = \text{rank}(H_p)$.

- In our example, $\beta_1 = \text{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- $\beta_p$ is the number of independent $p$-th holes.
- A tetrahedron has $\beta_0 = 1$ (connected), $\beta_1 = \beta_2 = 0$ (no holes or voids)
- A **hollow** tetrahedron has $\beta_0 = 1, \beta_1 = 0, \beta_2 = 1$
- Removing the four triangle faces, the edge skeleton has $\beta_0 = 1, \beta_1 = 3$ (one is the sum of the other three), $\beta_2 = 0$ (no more void).
- Removing the edges, $\beta_0 = 4$ (4 vertices) and $\beta_1 = \beta_2 = 0$. 

(Zhu, University of Wisconsin-Madison)
From data to simplicial complex

- Given data \( x_1, \ldots, x_n \in \mathbb{R}^d \).
From data to simplicial complex

- Given data \( x_1, \ldots, x_n \in \mathbb{R}^d \).
- If any subset of \( p + 1 \) points are within diameter \( \epsilon \), we add a \( p \)-simplex generated by those points.
From data to simplicial complex

- Given data $x_1, \ldots, x_n \in \mathbb{R}^d$.
- If any subset of $p + 1$ points are within diameter $\epsilon$, we add a $p$-simplex generated by those points.

**Definition**

A Vietoris-Rips complex of diameter $\epsilon$ is the simplicial complex $VR(\epsilon) = \{ \sigma \mid \text{diam}(\sigma) \leq \epsilon \}$. 

(Zhu, University of Wisconsin-Madison) Persistent homology
From data to simplicial complex

- Given data $x_1, \ldots, x_n \in \mathbb{R}^d$.
- If any subset of $p + 1$ points are within diameter $\epsilon$, we add a $p$-simplex generated by those points.

**Definition**

A *Vietoris-Rips complex of diameter* $\epsilon$ *is the simplicial complex* $VR(\epsilon) = \{\sigma \mid \text{diam}(\sigma) \leq \epsilon\}$.

**Example**

- VR(1)
- VR(2)
- VR($\sqrt{5}$)
Filtration

- Which $\epsilon$ should we pick?
Filtration

- Which $\epsilon$ should we pick?
- Don’t pick – look at all $\epsilon$’s
Filtration

- Which $\epsilon$ should we pick?
- Don’t pick – look at all $\epsilon$'s

Definition

An increasing sequence of $\epsilon$ produces a filtration, i.e., a sequence of increasing simplicial complexes $VR(\epsilon_1) \subseteq VR(\epsilon_2) \subseteq \ldots$, with the property that a simplex enters the sequence no earlier than all its faces.
Persistent homology

- In a filtration, at what value of $\epsilon$ does a hole appear, and how long does it persist till it is filled in?
Persistent homology

- In a filtration, at what value of $\epsilon$ does a hole appear, and how long does it persist till it is filled in?
- Barcode

![Diagram of barcode and VR complexes](image)

Zhu, University of Wisconsin-Madison
Applications to natural language processing

Good articles “tie back.”

How can we capture such loopy structure in text documents?
Applications to natural language processing

- Some documents “straight,” others “twist and turn”
Applications to natural language processing

- Some documents “straight,” others “twist and turn”
- Divide a document into small units $x_1, \ldots, x_n$ (e.g., sentences, paragraphs).
Applications to natural language processing

- Some documents “straight,” others “twist and turn”
- Divide a document into small units $x_1, \ldots, x_n$ (e.g., sentences, paragraphs).
- Given distance function $D(x_i, x_j) \geq 0$ (e.g., Euclidean, cosine)
Applications to natural language processing

- Some documents “straight,” others “twist and turn”
- Divide a document into small units $x_1, \ldots, x_n$ (e.g., sentences, paragraphs).
- Given distance function $D(x_i, x_j) \geq 0$ (e.g., Euclidean, cosine)
- We will focus on the 0-th (clusters) and 1st (holes) order homology classes.
Example: Itsy bitsy spider

The Itsy Bitsy Spider climbed up the water spout
Down came the rain and washed the spider out
Out came the sun and dried up all the rain
And the Itsy Bitsy Spider climbed up the spout again

- bag-of-words

```
<table>
<thead>
<tr>
<th>again</th>
<th>all</th>
<th>and</th>
<th>bitsy</th>
<th>came</th>
<th>climbed</th>
<th>down</th>
<th>dried</th>
<th>itsy</th>
<th>out</th>
<th>rain</th>
<th>spider</th>
<th>spout</th>
<th>sun</th>
<th>the</th>
<th>up</th>
<th>washed</th>
<th>water</th>
</tr>
</thead>
</table>
| 0     | 0   | 0   | 1     | 0    | 1       | 0    | 0     | 1    | 0   | 0    | 1      | 1     | 0   | 0   | 2   | 1      | 0     | 1
| 0     | 0   | 1   | 0     | 1    | 0       | 1    | 0     | 0    | 1   | 1    | 1      | 0     | 0   | 0   | 2   | 0      | 1     | 0
| 0     | 1   | 1   | 0     | 1    | 0       | 0    | 1     | 1    | 0   | 0    | 1      | 1     | 0   | 2   | 1   | 0      | 0     | 0
| 1     | 0   | 1   | 1     | 0    | 1       | 0    | 0     | 1    | 0   | 0    | 1      | 1     | 0   | 2   | 1   | 0      | 0     | 0 |
```
Example: Itsy bitsy spider

The Itsy Bitsy Spider climbed up the water spout
Down came the rain and washed the spider out
Out came the sun and dried up all the rain
And the Itsy Bitsy Spider climbed up the spout again

- bag-of-words

- vertices
Similarity Filtration (SIF)

\[ D_{max} = \max D(x_i, x_j), \forall i, j = 1 \ldots n \]

**FOR** \( m = 0, 1, \ldots M \)

Add \( VR \left( \frac{m}{M} D_{max} \right) \) to the filtration

**END**

Compute persistent homology on the filtration

- larger diameter, looser tie-backs
Similarity Filtration (SIF)

\[ D_{\text{max}} = \max D(x_i, x_j), \forall i, j = 1 \ldots n \]

FOR \( m = 0, 1, \ldots M \)

Add \( VR \left( \frac{m}{M} D_{\text{max}} \right) \) to the filtration

END

Compute persistent homology on the filtration

- larger diameter, looser tie-backs
- order of \( x_1 \ldots x_n \) ignored
Similarity Filtration with Time Skeleton (SIFTS)

\[
D(x_i, x_{i+1}) = 0 \text{ for } i = 1, \ldots, n - 1
\]
\[
D_{max} = \max D(x_i, x_j), \forall i, j = 1 \ldots n
\]

FOR \( m = 0, 1, \ldots M \)

Add \( VR \left( \frac{m}{M} D_{max} \right) \) to the filtration

END

Compute persistent homology on the filtration

- time edges allow tie-back in time
SIF vs. SIFTS on Itsy bitsy spider

SIF (dimension 0)

SIF (dimension 1)

SIFTS (dimension 0)

SIFTS (dimension 1)
On Nursery Rhymes and Other Stories

Row Row Row Your Boat

SIF (dimension 0)
SIF (dimension 1)

SIFTS (dimension 0)
SIFTS (dimension 1)

London Bridge

SIF (dimension 0)
SIF (dimension 1)

SIFTS (dimension 0)
SIFTS (dimension 1)

Little Red-Cap

Alice in Wonderland

- London Bridge: “My fair Lady” repeats 12 times.
On Nursery Rhymes and Other Stories

Row Row Row Your Boat

London Bridge

Little Red-Cap

- London Bridge: “My fair Lady” repeats 12 times.
- Little Red-Cap: “The better to see you with, my dear” and “The better to eat you with!”

[Zhu, University of Wisconsin-Madison]
On Child and Adolescent Writing

- Older writers have more complex barcodes?
On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8
On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8
- SIFTS barcode summary statistics:
On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8
- SIFTS barcode summary statistics:
  - holes?: what percentage of articles have $H_1$ holes
  - $\epsilon^*$: the smallest $\epsilon$ when the first hole in $H_1$ forms.

<table>
<thead>
<tr>
<th></th>
<th>Child</th>
<th>Adolescent</th>
</tr>
</thead>
<tbody>
<tr>
<td>holes?</td>
<td>87%</td>
<td>100%</td>
</tr>
<tr>
<td>$\epsilon^*$</td>
<td>1.35 ($\pm 0.02$)</td>
<td>1.27 ($\pm 0.02$)</td>
</tr>
</tbody>
</table>
On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8
- SIFTS barcode summary statistics:
  - holes?: what percentage of articles have $H_1$ holes
  - $|H_1|$: number of holes in the article

(Zhu, University of Wisconsin-Madison)
On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8
- SIFTS barcode summary statistics:
  - holes?: what percentage of articles have $H_1$ holes
  - $|H_1|$: number of holes in the article
  - $\epsilon^*$: the smallest $\epsilon$ when the first hole in $H_1$ forms.

<table>
<thead>
<tr>
<th></th>
<th>child</th>
<th>adolescent</th>
<th>adol. trunc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>holes?</td>
<td>87%</td>
<td>100%*</td>
<td>98%*</td>
</tr>
<tr>
<td>$</td>
<td>H_1</td>
<td>$</td>
<td>3.0 (±0.2)</td>
</tr>
<tr>
<td>$\epsilon^*$</td>
<td>1.35 (±.02)</td>
<td>1.27 (±.02)*</td>
<td>1.38 (±.01)</td>
</tr>
</tbody>
</table>

*: statistically significantly different from “child”
Is Homology Merely Counting Repeats?

- On $x_1 \leadsto x_2 \leadsto x_3$ where $x_1, x_2, x_3$ SIFTS will find two holes:
  
  \[ x_1 \leftrightarrow x_2, \ x_2 \leftrightarrow x_3 \]
Is Homology Merely Counting Repeats?

- On $x_1 \rightsquigarrow x_2 \rightsquigarrow x_3$ where $x_1, x_2, x_3$ SIFTS will find two holes:
  
  $x_1 \leftarrow x_2, x_2 \leftarrow x_3$

- $k$ such repeats of $x$ will generate $k - 1$ holes. The Betti number $\beta_1 = k - 1$?
Is Homology Merely Counting Repeats?

- On \( x_1 \rightsquigarrow x_2 \rightsquigarrow x_3 \) where \( x_1, x_2, x_3 \) SIFTS will find two holes:
  \[ x_1 \Rightarrow x_2, \ x_2 \Rightarrow x_3 \]
- \( k \) such repeats of \( x \) will generate \( k - 1 \) holes. The Betti number \( \beta_1 = k - 1 \)?
- No.

(Zhu, University of Wisconsin-Madison)
Is Homology Merely Counting Repeats?

- On $x_1 \rightsquigarrow x_2 \rightsquigarrow x_3$ where $x_1, x_2, x_3$ SIFTS will find two holes: $x_1 \implies x_2, x_2 \implies x_3$
- $k$ such repeats of $x$ will generate $k - 1$ holes. The Betti number $\beta_1 = k - 1$?
- No.

- $k - 1 = 3$, SIFTS correctly finds $\beta_1 = 1$
Is Homology Merely Counting Repeats?

- On $x_1 \leadsto x_2 \leadsto x_3$ where $x_1, x_2, x_3$ SIFTS will find two holes: $x_1 \mapsto x_2$, $x_2 \mapsto x_3$
- $k$ such repeats of $x$ will generate $k - 1$ holes. The Betti number $\beta_1 = k - 1$?
- No.

- Left: $k - 1 = 3$, SIFTS correctly finds $\beta_1 = 1$
- Right: $k - 1 = 12$, merging $x$ 0 holes, SIFTS correctly finds $\beta_1 = 2$
Persistent homology may offer new representations for machine learning.

To read more, see the references in Xiaojin Zhu. *Persistent homology: An introduction and a new text representation for natural language processing*. IJCAI, 2013.
Persistent homology may offer new representations for machine learning

How to best use it?

To read more, see the references in Xiaojin Zhu. **Persistent homology: An introduction and a new text representation for natural language processing.** IJCAI, 2013.